

第7章 数据库设计



沈明玉

合肥工业大学
Hefei University of Technology

计算机与信息学院

第7章 数据库设计

本章内容：

- 7.1 数据库设计概述
- 7.2 需求分析
- 7.3 概念结构设计
- 7.4 逻辑结构设计
- 7.5 物理结构设计
- 7.6 数据的实施和维护

合肥工业大学
Hefei University of Technology

计算机与信息学院

第7章 数据库设计

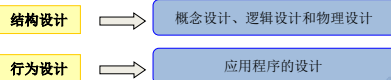
7.1 数据库设计概述

7.1.1 数据库设计的任务、内容和特点

■ 数据库设计的任务

数据库设计是指根据用户需求设计数据库结构的过程。

■ 数据库设计的内容



合肥工业大学

计算机与信息学院

7.1 数据库设计概述

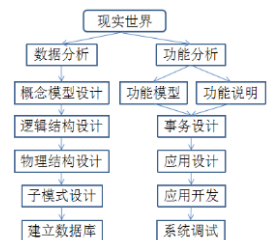
■ 数据库设计的特点

1) 数据库建设的基本规律

- 三分技术
- 七分管理
- 十二分基础数据

2) 结构（数据）设计和行为（处理）设计相结合

3) 反复性、试探性、分步进行



合肥工业大学

计算机与信息学院

7.1.2 数据库设计的基本步骤

1. 需求分析

获取用户需求，了解相关领域的业务知识，包括应用系统的应用环境和功能要求、具体业务处理方式等。

2. 概念结构设计

归纳与抽象需求分析阶段的分析结果，形成独立于具体DBMS的概念结构。

3. 逻辑结构设计

将概念结构转换为某个具体DBMS所支持的数据模型，并进行优化。

4. 物理结构设计

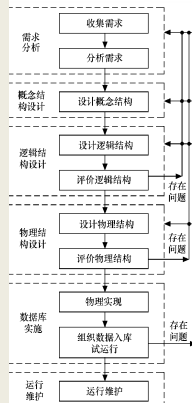
物理结构设计阶段为逻辑模型选取合适的物理结构。

5. 数据库实施

根据逻辑结构，创建数据库。编写应用程序和SQL程序，组织数据迁移，进行试运行。

6. 数据库运行和维护

数据库运行阶段需要不断维护，分析其性能，调整性能参数。对数据库的运行数据进行备份、恢复和处理等。



7.1 数据库设计概述

7.1.3 数据库设计与数据库的三级模式

◆ 需求分析阶段：综合用户的应用需求；

◆ 概念设计阶段：形成独立于DBMS的概念模式；

◆ 逻辑设计阶段：将概念模式（如：E-R图）转换成DBMS支持的数据模型（如：关系模型），形成数据库的逻辑模式；

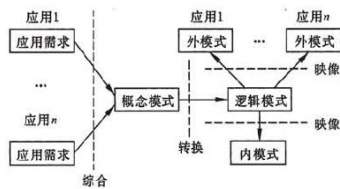
◆ 根据用户的处理要求和安全性的考虑：在基本表的基础上建立必要的视图，形成数据库的外模式；

◆ 物理设计阶段：根据DBMS的特点和处理的需要，选择存储结构，建立索引，形成数据库的内模式。

合肥工业大学

计算机与信息学院

7.1 数据库设计概述



第7章 数据库设计

7.2 需求分析

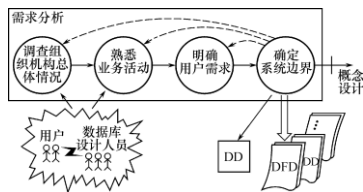
7.2.1 需求分析的任务

需求分析就是分析用户的需要与要求，确定系统必须完成哪些工作，对系统提出完整、准确、清晰、具体的要求。

- ◆ 调查现实世界要处理的对象（组织、部门、企业等）；
- ◆ 了解业务处理、处理的流程；
- ◆ 用户希望数据库应用程序的功能；
- ◆ 了解用户对系统的信息要求、处理要求以及安全性与完整性要求、处理的响应时间等。

7.2 需求分析

■ 需求分析过程



7.2 需求分析

7.2.2 需求分析的方法

- ① 跟班作业。亲自参与业务活动，了解业务处理的基本情况。
- ② 开调查会。通过与用户座谈、询问等方式来解决疑问。
- ③ 请专人介绍。
- ④ 设计调查表请用户填写。
- ⑤ 查阅记录。
- ⑥ 学习文件。
- ⑦ 使用旧系统。

7.2 需求分析

7.2.3 数据流图与数据字典

需求分析的阶段性结果以**数据流图**和**数据字典**的形式描述。

■ **数据流**是数据在系统内的传输途径，**数据流图**从数据传递和加工的角度，以图形的方式刻画数据流从输入到输出的变换过程。数据流图是结构化系统分析的主要工具，它去掉了具体的组织机构、工作场所、物流等，仅反映信息和数据存储、流动、使用以及加工的情况。

■ **数据字典**是各类数据描述的集合。通常包括数据项、数据结构、数据流、数据存储、处理过程和外部实体等6个部分。数据字典通过对数据项和数据结构的定义来描述数据流、数据存储的逻辑内容。

7.2 需求分析

■ 数据流图

- 用命名的箭头表示数据流
- 用圆圈表示处理
- 用矩形或其他形状表示存储



7.2 需求分析

■ 数据字典

内容：数据项、数据结构、数据流、数据存储、处理过程

● 数据项

数据项是不可再分的数据单位。

数据项描述=（数据项名，数据项含义说明，别名，数据类型，长度，取值范围，取值含义，与其他数据项的逻辑关系）；

● 数据结构

数据结构反映了数据之间的组合关系。

数据结构描述=（数据结构名，含义说明，组成：{数据项或数据结构}）；

合肥工业大学 计算机与信息学院

7.2 需求分析

● 数据流

数据流是数据结构在系统内传输的路径。

数据流描述=（数据流名，说明，数据流来源，数据流去向，组成：{数据结构}，平均流量，高峰期流量）；

● 数据存储

数据存储是数据结构停留或保存的地方，也是数据流的来源和去向之一。
数据存储描述=（数据存储名，说明，编号，流入的数据流，流出的数据流，

组成：{数据结构}，数据量，存取方式）

存取方法：批处理/联机处理；检索/更新；顺序检索/随机检索。

合肥工业大学 计算机与信息学院

7.2 需求分析

● 处理过程

处理过程的具体处理逻辑一般用判定表或判定树来描述。

数据字典中只需要描述处理过程的说明性信息。

处理过程描述=（处理过程名，说明，输入：{数据流}，输出：{数据流}，处理：{简要说明}）

简要说明：主要说明该处理过程的功能及处理要求。

✓ 功能：该处理过程用来做什么

✓ 处理要求：处理频度要求（如单位时间里处理多少事务，多少数据量）；响应时间要求等。

合肥工业大学 计算机与信息学院

第7章 数据库设计

7.3 概念结构设计

数据库的概念结构设计就是将需求分析得到的用户需求抽象为信息结构，即概念模型。

7.3.1 概念模型

概念模型是对信息世界的抽象表示。最常用的概念模型：**E-R模型**

■ 概念模型应具有的特点

- 能真实、充分地反映现实世界。各类数据及其相互之间的联系和约束；
- 易懂易用，能为非计算机专业人员所接受；
- 组成模型的概念少，定义严格，无多义性；
- 具有图形表示能力；
- 易于向其他数据模型转换。

7.3 概念结构设计

7.3.2 实体联系模型（E-R模型）

■ E-R模型设计方法

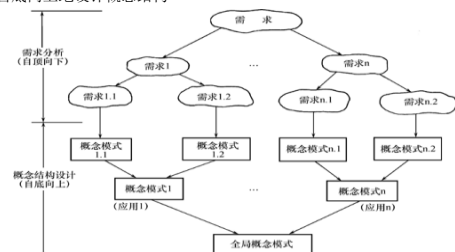


合肥工业大学 计算机与信息学院

7.3 概念结构设计

■ 常用策略

- 自顶向下地进行需求分析
- 自底向上地设计概念结构

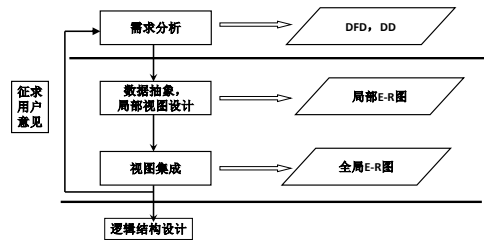


7.3 概念结构设计

■ 自底向上设计概念结构的步骤

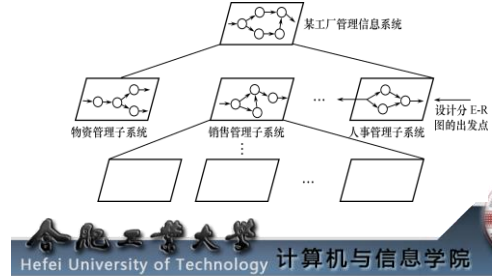
第1步：抽象数据并设计局部视图

第2步：集成局部视图，得到全局概念结构



7.3 概念结构设计

■ 选择局部应用



7.3 概念结构设计

■ 设计局部E-R图

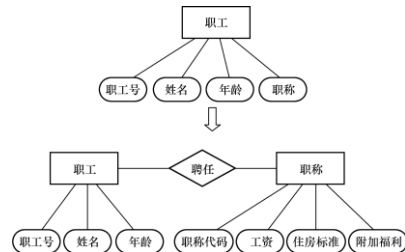
- 将各局部应用涉及的数据分别从数据字典中抽取出来；
- 参照数据流程图，标定各局部应用中的实体、实体的属性、标识实体的码；
- 确定实体之间的联系及其类型（1:1, 1:n, m:n）。

● 两条准则：

- （1）属性不能再具有需要描述的性质。即属性必须是不可分的数据项，不能再由另一些属性组成；
- （2）属性不能与其他实体具有联系。联系只发生在实体之间。

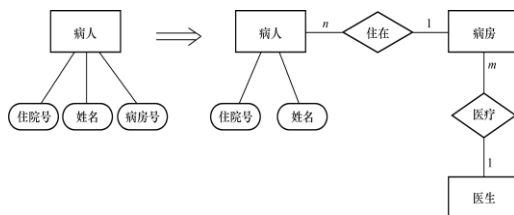
7.3 概念结构设计

■ 例：职工与职称



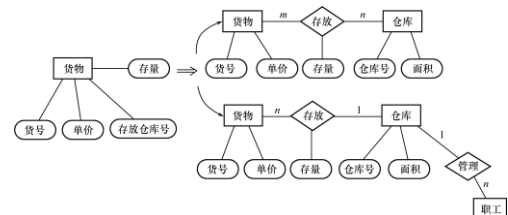
7.3 概念结构设计

■ 例：病人与病房



7.3 概念结构设计

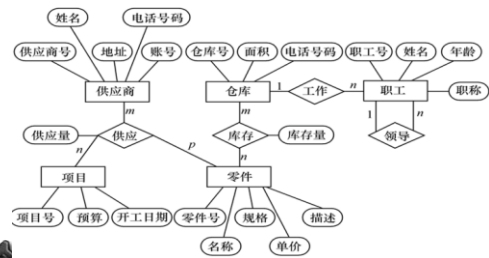
■ 例：货物与仓库



7.3 概念结构设计

■ 回顾：仓库管理E-R图

- 一个仓库可以存放多种零件，一种零件可以存放在多个仓库中。仓库和零件具有多对多的联系。用库存量表示某种零件在某个仓库中的数量。
- 一个仓库有多个职工当仓库保管员，一个职工只能在一个仓库工作，仓库和职工之间是一对多的联系。
- 职工之间具有领导-被领导关系。即仓库主任领导若干保管员。
- 供应商、项目和零件三者之间具有多对多的联系。



7.3 概念结构设计

■ 视图的集成

(1) 确认视图中的对应关系和冲突

对应关系：视图中语义都相同的概念，是视图的共同部分；

冲突：相互之间有矛盾的概念。

常见的冲突：命名冲突、概念冲突、域冲突、约束冲突

(2) 对视图进行修改，解决冲突

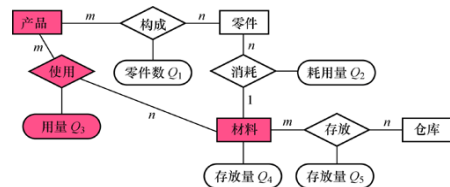
(3) 合并视图，形成全局视图。

合并对应的部分、保留特殊的部分、删除冗余部分，必要时做适当修改，力求视图简单清晰。

(4) 消除冗余

采用分析法，消除冗余的数据和联系。

■ 冗余的消除



第7章 数据库设计

7.4 逻辑结构设计

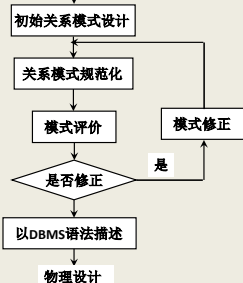
■ 任务

将概念设计阶段产生的数据库概念模式变换为数据库的逻辑模式。

■ E-R模型向关系模型转换

- ✓ 将实体转换为关系
- ✓ E-R图中联系的转换

概念结构设计



7.4 逻辑结构设计

■ E-R模型向关系模型转换

□ 实体映射为关系

一个实体映射一个关系（模式），实体的常规属性成为关系的一列（属性），若实体中存在单值的唯一属性即为关系的主码。

■ E-R图中联系的转换

□ 1:M 联系的转换

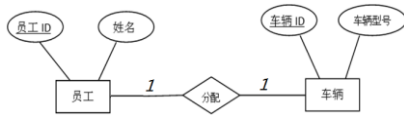
M端实体对应的关系中增加一个外码，该外码是1端实体对应关系的主码。

□ M:N 联系的转换

联系两端的主码属性与联系自身的属性构成一个新的关系，该关系有两个外码，对应多对多联系中两个实体的主码，这两个外码构成了新关系的复合主码。

7.4 逻辑结构设计

□ 1:1 联系的转换



7.4 逻辑结构设计

转换方法与 1:M 联系类似。

两种情况均可，即：

- 员工关系中加入车辆ID外码；
- 车辆关系中加入员工ID外码。

但如果两个外码有一个是强制的，而另一个是可选的（如上例），则选择强制的外码（车辆关系中的员工ID）效果更好。

还应考虑应用的实际情况！

7.4 逻辑结构设计

■ 逻辑结构的优化

完成E-R图向关系数据模型的转换之后，还需要对数据模型进行优化，修改、调整数据模型的结构，提高数据库的性能。

□ 关系模式的规范化

- ① 确定数据依赖。按需求分析阶段得到的语义，分别写出每个关系模式内部各属性之间的数据依赖以及不同关系模式属性之间的数据依赖。
- ② 对于各个关系模式之间的数据依赖进行极小化处理，消除冗余的联系。
- ③ 按照数据依赖的理论对关系模式逐一进行分析，考查是否存在部分函数依赖、传递函数依赖、多值依赖等，确定各关系模式分别属于第几范式。
- ④ 按照需求分析阶段得到的信息要求和处理要求，分析这些模式是否满足这些要求，确定是否要对某些模式进行合并或分解。

7.4 逻辑结构设计

□ 性能优化

- 结合需求分析中的性能需求（反规范化设计）。
- 并不是规范化程度越高的关系就越优。当一个应用的查询中经常涉及两个或多个关系模式的属性时，系统必须经常地进行连接运算，而连接运算的代价是相当高的，因此在这种情况下，2NF甚至1NF也许是最好的。
- 关系模式的分解：水平分解、垂直分解。
- 关系模式的合并等。

7.4 逻辑结构设计

■ 用户视图-局部逻辑结构的设计

（1）使用更符合用户习惯的别名

视图集成时，为了减少异义同名的冲突，规范了一些名称。规范后的名称与局部用户的习惯不一致，将影响用户的工作。设计用户的子模式时可以重新定义某些属性名，使其与用户习惯一致。

（2）针对不同级别的用户定义不同的视图

不同的用户关心的实体及其属性不同，对这些实体与属性的访问权限也不同。例如学生允许查看课程的开设情况和任课教师的学历、职称等属性，但是没有权利查看教师的籍贯、工资等属性。为了满足系统安全性要求，需要针对不同级别的用户定义不同的视图。

7.4 逻辑结构设计

（3）简化用户对系统的使用

某些局部应用中经常要使用一些很复杂的查询，为了方便用户，可以将这些复杂查询定义为视图，用户每次只对定义好的视图进行查询，方便用户使用系统。

（4）提高数据的逻辑独立性

对于数据库应用系统的开发者尤为重要。

第7章 数据库设计

7.5 物理结构设计

- 数据库是存储在物理设备上的，数据库在物理设备上的存储结构与存取方法称为数据库的物理结构，物理结构依赖于给定的DBMS和计算机系统。
- 逻辑数据库设计工作完成后，需要为给定的逻辑数据模型选取一个适合应用环境的物理结构，对物理结构进行时间（执行效率）和空间效率的评价。
- 数据库物理设计的任务是对给定的逻辑数据模型选取适合应用环境的物理结构，即在逻辑设计的基础上，为每个关系模式选择合适的存储结构和存取方法，使数据库的事务能够高效率地运行。

7.5 物理结构设计

■ 数据库物理设计阶段主要包括以下4个过程：

- ① 分析影响数据库物理结构设计的因素。
- ② 为关系模式选择存取方法。
- ③ 设计关系、索引等数据库文件的物理存储结构。
- ④ 评价物理结构。

Hefei University of Technology 计算机与信息学院

7.5 物理结构设计

7.5.1 影响物理结构设计的因素

- ① 与数据库查询事务有关的因素
 - ✓ 要查询的关系；
 - ✓ 查询条件所涉及的属性；
 - ✓ 连接条件所涉及的属性；
 - ✓ 查询的投影属性。
- ② 与数据库更新事务有关的因素
 - ✓ 要更新的关系；
 - ✓ 每个关系上更新操作的类型；
 - ✓ 删除和修改操作条件所涉及的属性；
 - ✓ 修改操作要改变的属性值。
- ③ 每个事务在各个关系上运行的频率和时间约束

Hefei University of Technology 计算机与信息学院

7.5 物理结构设计

7.5.2 关系模式存取方法选择

目的：使事务能快速存取数据库中的数据。

DBMS一般都提供多种存取方法，如：索引(index)、Hash、聚簇(cluster)。

1. B+树索引存取方法的选择

根据实际需要确定对哪些关系的哪些属性列建立索引、组合索引或唯一索引。是否需要建立索引，可以考虑以下原则。

- ① 如果一个（组）属性经常作为查询条件，可以考虑建立索引（组合索引）。
- ② 如果一个（组）属性经常使用聚集函数，可以考虑建立索引。
- ③ 如果一个（组）属性经常需要作为连接条件，可以考虑建立索引。

◆ 注意：索引的利弊！

Hefei University of Technology 计算机与信息学院

7.5 物理结构设计

2. Hash索引存取方法的选择

Hash方法是用Hash函数存储和存取关系元组的方法。指定某个关系上的一个（组）属性A作为Hash码，为该Hash码定义一个函数（称为Hash函数），关系元组的存储地址由Hash(a)决定，a是该元组在属性A上的值。

如果一个关系的属性主要出现在等值连接条件中或主要出现在等值比较的选择条件中，是否选择Hash存取方法的原则如下：

- ① 一个关系的大小可预知，而且不变，则此关系可以选择Hash存取方法。
- ② 如果DBMS支持动态Hash存取方式，即使关系大小动态可变，也可以采用Hash存取方法。

7.5 物理结构设计

3. 聚簇存取方法的选择

为了提高某个属性（或属性组）的查询速度，把这个或这些属性（称为**聚簇码**）上具有相同值的元组集中存放在连续的物理块上，这种方法称为**聚簇**。

聚簇方法可以大大提高按聚簇码进行查询的效率。

■ 需要建立多少个聚簇，每个聚簇中包括哪些关系，可以参考以下原则：

- ① 经常在一起进行连接操作的关系可以建立聚簇。
- ② 如果一个关系的一组属性经常出现在相等比较条件中，则该单个关系可以建立聚簇。
- ③ 如果一个关系的一个（组）属性上的值重复率很高，则此关系可以建立聚簇。
- ④ 一个数据库可以建立多个聚簇，一个关系只能加入一个聚簇。

7.5 物理结构设计

■ 如果出现下列情况，建议不要建立聚簇：

- ① 经常需要进行全表扫描的关系。
- ② 更新操作频率高于连接操作的关系。

■ 由于一个关系只能有一个聚簇，如何运用需要根据应用特点进行分析，以得到较好的聚簇方案。

合肥工业大学

计算机与信息学院



7.5 物理结构设计

7.5.3 确定数据库的存储结构

✓ 目的：确定如何在磁盘上存储关系、索引等数据库文件，使得空间利用率最大而数据库操作的开销最小。（不同的DBMS平台差异较大！）

✓ 综合考虑：存取时间、空间利用率和维护代价三方面因素。

1. 确定数据的存放位置

- 应将易变部分与稳定部分、经常存取部分与存取频率较低部分分开存放；
- 减少访盘冲突，提高I/O的并行性；
- 分散热点数据，均衡I/O负载；
- 保证关键数据的快速访问，缓解系统的瓶颈。

合肥工业大学

计算机与信息学院



7.5 物理结构设计

2. 确定系统配置

- DBMS会提供一些系统配置变量和存储分配参数，供物理优化使用。
- 初始情况下，系统都为这些变量赋予了合理的缺省值。可以进行针对性调整。
- 系统配置变量包括：同时使用数据库的用户数，同时打开的数据库对象数，使用的缓冲区长度、个数，时间片大小、数据库的大小，装填因子，锁的数目等，这些参数值影响存取时间和存储空间分配，在物理设计时就要根据应用环境确定这些参数值，以使系统性能最优。
- 物理设计时的调整只是初步的，系统运行时还要根据系统实际运行情况做进一步的调整，以期切实改进系统性能。

合肥工业大学

计算机与信息学院



7.5 物理结构设计

3. 评价物理结构

物理设计过程中需考虑时间和空间效率、维护代价和用户的要求等，对这些因素考虑的侧重点不同会产生多种物理设计方案。

因此，数据库的物理结构应全面权衡这几个因素，对各种可能的设计方案进行评价，评价的重点是系统的时间和空间效率，并从多个方案中选出较优的物理结构。

合肥工业大学

计算机与信息学院

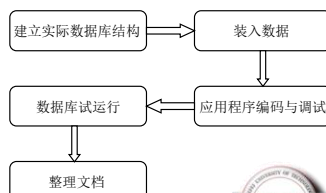


第7章 数据库设计

7.6 数据库的实施和维护

7.6.1 数据库的实施

数据库实施是指根据逻辑设计和物理设计的结果，在计算机上建立起实际的数据库结构、装入数据、进行测试和试运行的过程。



合肥工业大学

计算机与信息学院



7.6 数据库的实施和维护

7.6.2 数据库的运行和维护

数据库运行和维护阶段的主要任务包括以下三项内容：

- (1) 维护数据库的安全性及完整性；
- (2) 监测并改善数据库性能；
- (3) 重新组织和构造数据库。

◆ 只要数据库系统在运行，就需要不断地进行修改、调整和维护！

合肥工业大学

计算机与信息学院





■ 本章思考题:

数据库概念设计是必须的吗? 如何理解其存在的价值?

■ 本章作业:

习题8、10

